Canadian Agency for Drugs and Technologies in Health

RAPID RESPONSE REPORT: SUMMARY WITH CRITICAL APPRAISAL

TITLE: Audiograms and Functional Auditory Testing to Assess Hearing Speech in Noise: A Review of the Clinical Evidence

DATE: 17 August 2015

CONTEXT AND POLICY ISSUES

Auditory fitness for duty is defined as "the possession of hearing abilities sufficient for safe and effective job performance".[1] Many occupations require sufficient hearing ability to be able to effectively perform duties, and as such, the ability to accurately assess hearing is vital.[1]

There are a number of ways to assess hearing ability. Pure-tone audiometry is most commonly used to assess hearing, and involves the administration of pure-tones at different intensities and frequencies to assess the softest sound audible to the individual.[1,2] An audiogram is then created based on the results of the audiometry to examine the hearing level in decibels (dB) at the different frequencies tested, measured in hertz (Hz).[2] Individual results are then compared to recommended ranges to assess for the presence and severity of hearing loss.[2]

While pure-tone audiometry is the most commonly used test to assess hearing, particularly in the context of auditory fitness for duty, it is not without limitations.[1] Cutoffs for the definition of hearing loss vary, and perhaps most important for assessing auditory fitness for duty, pure-tone audiometry is designed to be completed in an environment without background noise, which may not reflect the actual working environment.[1,2] As a result, while a person may have an audiogram that indicates normal hearing ability, they may actually have reductions in hearing ability when background noise is present.[1] Based on the limitations of pure-tone audiometry, functional auditory tests are suggested to measure the ability to hear in the presence of background noise. Such tests include the Hearing in Noise Test (HINT), the Speech Perception in Noise Test (SPIN), the Speech Recognition in Noise Test (SPRINT), and the Words in Noise Test (WIN). Each of these tests involves the participants listening to words or sentences with background noise present, to assess the ability of the participant to hear the words or sentences being spoken.[1] Further detail on each of these tests is available in Appendix 1.

The purpose of this Rapid Response report is to review the effectiveness of audiograms and functional auditory tests for assessing the ability to hear speech in noise.

RESEARCH QUESTIONS

1. What is the effectiveness of audiograms alone to assess the ability to hear speech in noise?

2. What is the comparative effectiveness of audiograms compared with functional tests for assessing the ability to hear speech in noise?

3. What is the comparative effectiveness of different functional tests for assessing the ability to hear speech in noise?

KEY FINDINGS

Audiograms were found to be related to the Words in Noise Test (WIN), but not the Hearing in Noise Test (HINT). The performance of the WIN and the Speech Recognition in Noise Test (SPRINT) were found to be similar in one study, and the WIN was found to be a more sensitive measure of speech-in-noise relative to the HINT in another study.

METHODS

Literature Search Methods

A limited literature search was conducted on key resources including PubMed, The Cochrane Library, University of York Centre for Reviews and Dissemination (CRD) databases, ECRI, Canadian and major international health technology agencies, as well as a focused Internet search. Methodological filters were applied to limit retrieval to health technology assessments, systematic reviews, meta-analyses, randomized controlled trials, non-randomized studies, and diagnostic studies. Where possible, retrieval was limited to the human population. The search was also limited to English language documents published between January 1, 2005 and July 17, 2015.

Rapid Response reports are organized so that the evidence for each research question is presented separately.

Selection Criteria and Methods

One reviewer screened citations and selected studies. In the first level of screening, titles and abstracts were reviewed and potentially relevant articles were retrieved and assessed for inclusion. The final selection of full-text articles was based on the inclusion criteria presented in Table 1.

Table 1: Selection Criteria	
Population	Adults
Intervention	Questions 1 and 2: Audiograms Question 3: Functional auditory tests (Hearing in Noise Test [HINT], Speech Perception in Noise Test [SPIN], Speech Recognition in Noise Test [SPRINT], SPRINT 100 (subscript), Words in Noise Test [WIN])

Table 1: Selection Criteria	
Comparator	Question 1: Not applicable Questions 2 and 3: Functional auditory tests (Hearing in Noise Test [HINT], Speech Perception in Noise Test [SPIN], Speech Recognition in Noise Test [SPRINT]
Outcomes	Effectiveness to determine the ability to hear speech in noise
Study Designs	Health technology assessments, systematic reviews, meta-analyses, randomized controlled trials, and non-randomized studies.

Exclusion Criteria

Articles were excluded if they did not meet the selection criteria outlined in Table 1, they were duplicate publications, or were published prior to 2005.

Critical Appraisal of Individual Studies

The included non-randomized studies were appraised using the Downs and Black checklist.[3] Summary scores were not calculated for the included studies; rather, a review of the strengths and limitations of each included study were described.

SUMMARY OF EVIDENCE

Details of study characteristics, critical appraisal, and study findings are located in Appendices 3, 4, and 5, respectively.

Quantity of Research Available

A total of 627 citations were identified in the literature search. Following screening of titles and abstracts, 607 citations were excluded and 20 potentially relevant reports from the electronic search were retrieved for full-text review. Thirty-seven potentially relevant publications were retrieved from the grey literature search. Of these potentially relevant articles, 53 publications were excluded for various reasons, while four publications met the inclusion criteria and were included in this report. Appendix 2 describes the PRISMA flowchart of the study selection.

Summary of Study Characteristics

Study Design

All studies used an observational repeated measures design, where both the intervention and comparison tests were administered to each participant, to evaluate the study objectives.[4-7]

Country of Origin

The included studies originated from the United States of America.[4-7]

Patient Population

The total number of study participants ranged from 72 individuals to 3430 individuals, and all were subjects obtained from a clinical setting.[4-7] Three of the studies did not report the proportion of males and females.[4-6] Two of the studies assessed younger individuals (mean age 23.3 years and 24.9 years, respectively) with normal hearing and older individuals (mean age 69.9 years and 70.8 years, respectively) with hearing loss.[6,7] Another study included 215 individuals with a mean age of 33 years.[4] The last study included 3430 veterans with a mean age of 62.3 years.

There were participants with hearing loss included in all studies, but the definition of hearing loss varied.[4-7] In particular, one study excluded those with pure-tone measurements > 40 decibels hearing loss (dB HL),[6] two studies excluded those with pure-tone measurements > 60 dB HL,[5,7] and the last study included individuals regardless of their pure-tone measurements.[4]

Interventions and Comparators

Two of the studies compared pure-tone audiograms to functional tests, including the HINT and the WIN.[4,5] The other two studies compared different functional tests; the first comparing the SPRINT to the WIN,[6] and the second comparing the HINT to the WIN.[7]

Outcomes

Outcomes included evaluating the relationship between audiogram high-frequency pure-tone average and scoring on the HINT and WIN,[4,5] assessing the validity of the SPRINT relative to the WIN for recognition performance,[6] and comparing the between-group differences in tests scores and pure-tone audiogram results with the HINT and the WIN.[7] All studies evaluated outcomes in groups stratified by hearing ability.[4-7]

Summary of Critical Appraisal

In terms of study strengths, all studies clearly described the study objectives and the tests and procedures used to administer the tests in each study.[4-7] All tests were administered in one sitting to ensure the tests were measuring hearing at one point in time.[4-7]

There were a number of limitations associated with the included studies. Only one study reported proportion of males and females in the study population.[7] In addition, investigators were not blinded to the tests in any of the included studies, and it is unclear if it was the investigators actually administering the tests and calculating the scores for each participant.[4-7] This is of particular concern relating to the studies assessing the WIN because the first author of these three studies was the creator of the WIN.[5-7] Sample size calculations were not performed for any of the studies, and as a result, for studies that did not find a statistical difference between groups, it was unclear whether this was due to their being truly no difference, or due to type 2 error. Also, all studies reported continuous measurements using mean and standard deviation, but did not report assessing the continuous variables for normal distribution.[4-7] Lastly, two of the studies did not report their statistical analysis procedures nor did they report how study outcomes were measured.[5,6]

Summary of Findings

What is the effectiveness of audiograms alone to assess the ability to hear speech in noise?

No studies were identified examining the effectiveness of audiograms to assess the ability to hear speech in noise.

What is the comparative effectiveness of audiograms compared with functional tests for assessing the ability to hear speech in noise?

Two studies compared pure-tone audiograms to functional tests for the ability to assess speech in noise. The first compared audiometry to the HINT,[4] and the second compared audiometry to WIN.[5]

The first study comparing audiometry to HINT used high-frequency pure-tone thresholds to stratify individuals into categories of hearing: normal hearing was defined as 10 – 15 dB HL, slight hearing loss was defined as 20 – 25 dB HL, mild hearing loss was defined as 30 – 40 dB HL, moderate hearing loss was defined as 45 – 55 dB HL, severe hearing loss was defined as 60 – 70 dB HL, and profound hearing loss was defined as 75 – 95 dB HL.[4] All individuals also underwent testing with the HINT, and composite scores were calculated. Scatterplots of the HINT composite scores stratified by the categories of hearing demonstrated similar performance across pure-tone threshold groups.[4] In addition, no significant difference was found in HINT composite scores for all hearing categories relative to the normal hearing category, except for the comparison with the profound hearing loss group.[4] The study authors concluded that "Results from the present study are consistent with investigations that have shown a poor relationship between pure-tone thresholds and speech recognition in noise performance" and "[t]he audiogram is not useful to infer the ability to recognize speech in noise."[4]

The second study compared high-frequency pure-tone thresholds to the WIN.[5] The audiogram was conducted on both the left and right ears, at frequencies of 250, 500, 1000, 2000, 3000, 4000, and 8000 Hz.[5] Individuals then underwent the WIN, with the multitalker babble dB sound pressure level (SPL) varying based on the pure-tone average obtained from the audiogram (those with a pure-tone average of ≤ 40 dB HL had multitalker babble presented at 80 dB SPL, and those with a pure-tone average of 42 – 58 had multitalker babble presented at 90 dB SPL).[5] When the WIN 50% correct point (measured as decibels signal-to-noise ratio) were plotted with the high-frequency pure-tone average, the bivariate plot indicated a clear linear relationship between the two tests, with an R^2 of 0.57 and a slope of 2.4 dB/dB.[5] Of note, performance on the WIN was significantly better in the right ear compared with the left ear, however, this was attributed to using the left ear first to test the WIN.[5] The authors concluded that, based on the results of the study, the WIN provides an accurate way of measuring word recognition in background noise, based on the assumption that audiograms effectively measure hearing loss related to speech in noise.[5]

What is the comparative effectiveness of different functional tests for assessing the ability to hear speech in noise?

Two publications assessed the effectiveness between different functional tests. One study compared the SPRINT to the WIN,[6] and the second compared the HINT to the WIN.[7]

The SPRINT was compared to the WIN in 72 people, 24 of which were young listeners with normal hearing (defined as ≤ 20 dB HL at the 250 – 8000 Hz octave intervals on audiogram), and 48 of which were older individuals (60 to 85 years of age) with hearing loss.[6] Hearing loss was defined as audiogram thresholds of 15 – 30 dB HL at 500 Hz, 20 – 40 dB HL at 1000 Hz, a three frequency pure-tone average of ≤ 40 dB HL at 500, 1000, and 2000 Hz, word-recognition scores in quiet of ≥ 40%, and no history of retrocochlear or middle ear pathology.[6] When a bivariate plot was created of the 50% correct point on the WIN and the percent correct on the

SPRINT of all study participants (including those with normal hearing and those with hearing loss identified on the audiogram), there was a clear linear relationship with the dB signal-to-noise ratio (S/N) on the WIN increasing as the percent correct on the SPRINT decreased.[6] The Pearson product-moment correlation for the performance between the SPRINT and the WIN was $r = -0.81$, $P < 0.01$.[6] The study authors concluded that "Graphically and numerically the SPRINT and WIN were highly related, which is indicative of good concurrent validity of the SPRINT."[6]

The second study compared the HINT with the WIN in 96 participants.[7] Similar to the study that compared the SPRINT and the WIN, participants were stratified into two groups: the first was a group of 24 young listeners with normal hearing, defined as defined as ≤ 20 dB HL, and the second was a group of 72 older listeners with hearing loss.[7] In this study, hearing loss was defined as pure-tone averages between 20 – 60 dB HL measured at 500, 1000, and 2000 Hz.[7] All participants underwent the WIN and the HINT, and 50% correct scores for each test were similar to what has been previously reported in the literature.[7] When a bivariate plot was created, the comparison of the HINT and WIN demonstrated that almost all of the 50% correct points for the WIN had a better signal-to-noise ratio compared with the HINT 50% correct points, indicating that the WIN was a more sensitive measure for speech-in-noise compared to the HINT.[7] The authors concluded that the WIN provided a more accurate assessment of the recognition performance between the normal hearing and hearing loss groups than did the HINT.[7]

Limitations

All of the study participants were identified through a clinical setting; therefore it is unclear whether the results of this study would be generalizable to assessing auditory fitness for duty. In addition, the included studies had a number of serious limitations, impacting the validity of the conclusions drawn from the studies. In particular, the three studies evaluating the WIN were conducted by the creator of the WIN, and it does not appear that the author was blinded to the test results during the study analysis.[5-7] Also, for the two studies comparing functional tests, a pure-tone audiogram was first used to stratify people into normal hearing and hearing loss groups, and the functional tests were then evaluated on their ability to distinguish between these two groups. As a result, the conclusions of these studies must be interpreted with caution given that it is unclear whether audiograms are truly effective at measuring speech in noise. In addition, definitions of hearing loss varied across studies, and study populations were not clearly described, particularly in the context of proportions of males and females. There were no studies identified for this review that included Speech Perception in Noise (SPIN) test, therefore no conclusions can be drawn about this test. Lastly, it is unclear whether the outcomes used in the included studies would translate into benefit in terms of job performance.

CONCLUSIONS AND IMPLICATIONS FOR DECISION OR POLICY MAKING

Based on the results of the four included studies, audiograms were found to be related to the WIN, but not the HINT. The performance of the WIN and the SPRINT were found to be similar in one study, and the WIN was found to be a more sensitive measure of speech-in-noise relative to the HINT in another study. The conclusions of this Rapid Response report are limited by the poor quality of the included studies, and it is unclear whether the outcomes reported in the included studies are applicable for assessing auditory fitness for duty. Studies are needed in the auditory fitness for duty population to assess the validity of tests for detecting speech in noise.

PREPARED BY:
Canadian Agency for Drugs and Technologies in Health
Tel: 1-866-898-8439
www.cadth.ca

REFERENCES

1. Tufts JB, Vasil KA, Briggs S. Auditory fitness for duty: a review. J Am Acad Audiol. 2009 Oct;20(9):539-57.

2. Walker JJ, Cleveland LM, Davis JL, Seales JS. Audiometry screening and interpretation. Am Fam Physician. 2013 Jan 1;87(1):41-7.

3. Downs SH, Black N. The feasibility of creating a checklist for the assessment of the methodological quality both of randomised and non-randomised studies of health care interventions. J Epidemiol Community Health [Internet]. 1998 Jun [cited 2015 Aug 17];52(6):377-84. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1756728/pdf/v052p00377.pdf

4. Vermiglio AJ, Soli SD, Freed DJ, Fisher LM. The relationship between high-frequency pure-tone hearing loss, hearing in noise test (HINT) thresholds, and the articulation index. J Am Acad Audiol. 2012 Nov;23(10):779-88.

5. Wilson RH. Clinical experience with the words-in-noise test on 3430 veterans: comparisons with pure-tone thresholds and word recognition in quiet. J Am Acad Audiol. 2011 Jul;22(7):405-23.

6. Wilson RH, Cates WB. A comparison of two word-recognition tasks in multitalker babble: Speech Recognition in Noise Test (SPRINT) and Words-in-Noise Test (WIN). J Am Acad Audiol. 2008 Jul;19(7):548-56.

7. Wilson RH, McArdle RA, Smith SL. An Evaluation of the BKB-SIN, HINT, QuickSIN, and WIN Materials on Listeners With Normal Hearing and Listeners With Hearing Loss. J Speech Lang Hear Res. 2007 Aug;50(4):844-56.

8. Nilsson M, Soli SD, Sullivan JA. Development of the Hearing in Noise Test for the measurement of speech reception thresholds in quiet and in noise. J Acoust Soc Am. 1994 Feb;95(2):1085-99.

9. Cord MT, Walden BE, Atack RM. SPeech Recognition In Noise Test (SPRINT) for H-3 profile [Internet]. Bethesda (MD): Walter Reed Army Medical Center; Military Audiology Association; [2009?]. [cited 2015 Aug 17]. Available from: http://militaryaudiology.org/site/wp-content/downloads/SPRINT/sprint_instructions_handout.doc

10. Wilson RH. Development of a speech-in-multitalker-babble paradigm to assess word-recognition performance. J Am Acad Audiol. 2003 Nov;14(9):453-70.

APPENDIX 1: Functional Tests to Assess Speech in Noise

The Hearing in Noise Test (HINT)

The HINT is an adaptive test that is composed of 250 sentences that are divided into 25 lists.[8] The test is adaptive in that the signal-to-noise ratio is adjusted based on the performance of the participant.[8] Only the speech level is adjusted whereas the noise level remains constant.[8] For each correct sentence the participant identifies, the following sentence is presented at a lower speech level, and for each sentence incorrectly identified, the following sentence is presented at a higher speech level.[8] The intention of the adaptive test is to ensure each participant approaches the 50% correct response rate.[8] The speech level typically is presented at 55 dB SPL to start, and the noise remains constant at 65 dB SPL.[8] The test is designed to be administered with the participant sitting 1 meter from eight speakers.[8] The speaker directly in front of the individual will present the sentences, whereas the other 7 speakers surrounding the participant will play the background noise.[8]

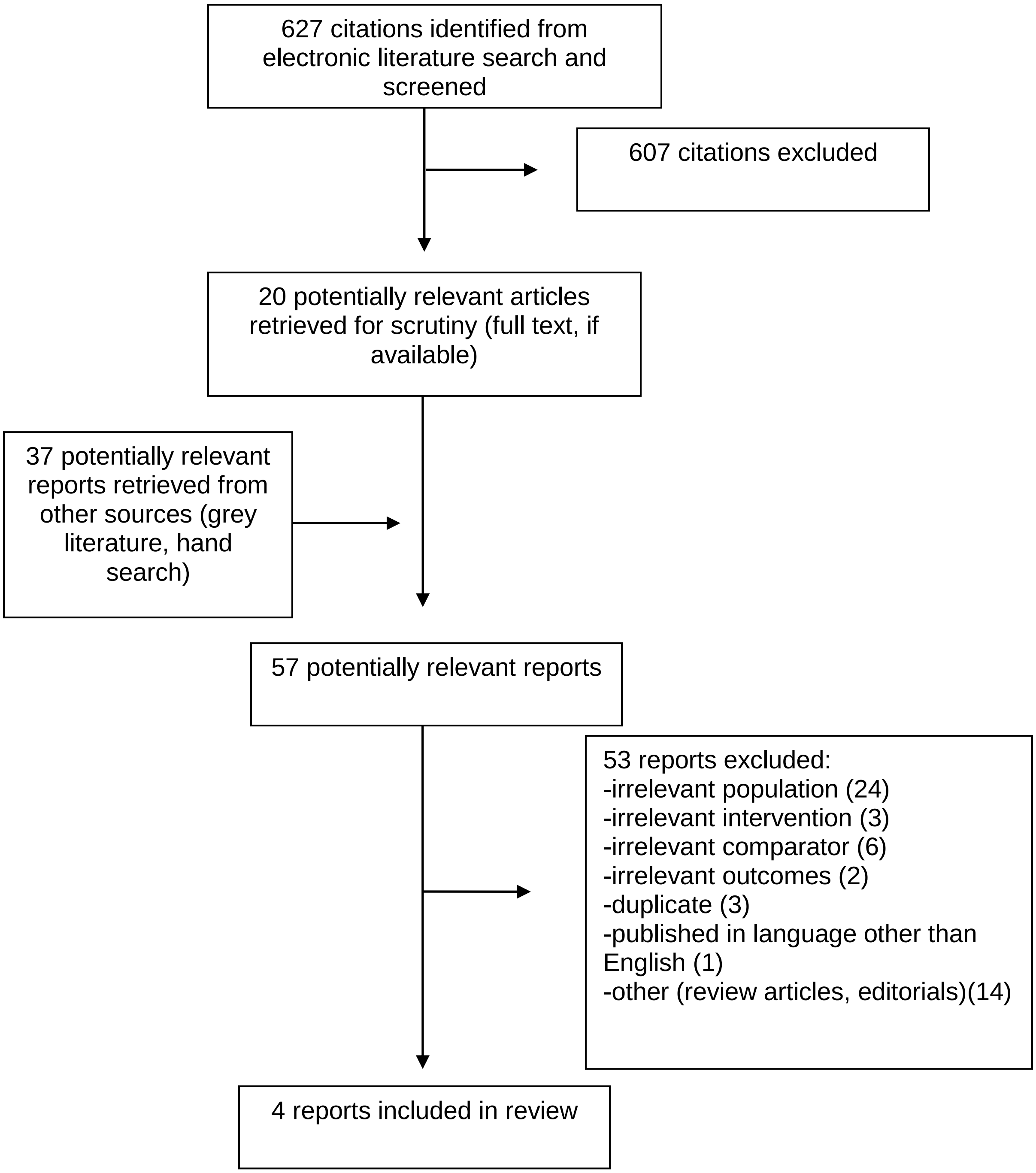
The Speech Recognition in Noise Test (SPRINT)

The SPRINT was designed by the United States Army to identify hearing loss in active duty soldiers.[9] The test includes 200 monosyllabic words that are pre-recorded with multitalker babble, and are delivered to both ears simultaneously using earphones.[9] The speech-to-babble ratio is 9dB; at this ratio, active duty soldiers with normal hearing are expected to hear at least 95% of the words correctly.[9] The SPRINT is recommended to be administered by an audiologist or a technician with audiologist supervision.[9]

The Words in Noise Test (WIN)

The WIN consists of the administration of 70 monosyllabic words divided into two 35 word lists that are pre-recorded with a noisy background.[10] The test is adaptive in that the loudness of the speech fluctuates during the test while the multitalker babble level remains constant.[10] The test is administered using earphones and is conducted in each ear separately.[10] The 50% correct point, measured in dB signal-to-babble ratio is then calculated for the participant.[10]

APPENDIX 2: Selection of Included Studies



APPENDIX 3: Characteristics of Included Publications

Table A1: Characteristics of Included Clinical Studies

First Author, Publication Year, Country	Study Design	Patient Characteristics	Intervention(s)	Comparator(s)	Clinical Outcomes
Vermiglio, 2012, USA[4]	Observational repeated measures design	n = 215 included of 278 tested Mean age = 33 years (range 17 – 59 years) All participants were native speakers of American English n = 63 excluded, including those with conductive hearing loss, those with asymmetrical hearing loss, those with non-response to any frequency, and those in threshold groups that were > +/- 2 standard deviations around the mean	High-frequency pure-tone audiogram, divided into categories: Normal -10 – 15 dB HL (n = 51) Slight 20 – 25 dB HL (n = 56) Mild 30 – 40 dB HL (n = 63) Moderate 45 – 55 dB HL (n = 29) Severe 60 – 70 dB HL (n = 12) Profound 75 – 95 dB HL (n = 4) Collected using the Hughson-Weslake procedure	HINT score, calculated with the formula: ([2 X Noise Front] + Noise Right + Noise Left)/4 Collected using a Windows-based HINT software system	• Relationship between the pure tone average and HINT thresholds in quiet and noise • Differences in HINT performance between pure tone threshold groups
Wilson, 2011, USA[5]	Observational repeated measures design	n = 3430 veterans with pure-tone audiograms, although some individuals did not undergo the WIN, resulting in 3291 being analyzed Mean age = 62.3 years (SD	Pure-tone audiogram, collected in the left and right ears, at frequencies of 250, 500, 1000, 2000, 3000, 4000, and 8000 Hz Collected using the modified	WIN, evaluated in the left and right ears separately Included 70 words, spoken by a female speaker Multitalker babble presented at 80	• To evaluate the relationship between the pure tone average and WIN thresholds in quiet and noise

Table A1: Characteristics of Included Clinical Studies					
First Author, Publication Year, Country	**Study Design**	**Patient Characteristics**	**Intervention(s)**	**Comparator(s)**	**Clinical Outcomes**
		= 12.8 years) Excluded those with a pure-tone average of ≥ 60 dB HL	Hughson-Weslake procedure	dB SPL when pure-tone average was ≤ 40 dB HL and 90 dB SPL when pure-tone average was 42 – 58 dB HL	
Wilson, 2008, USA[6]	Observational repeated measures design	n = 72, divided into: n = 24 (mean age = 23.3 years, range 19 – 29 years) with normal hearing, defined as ≤ 20 dB HL at the 250 – 8000 octave intervals n = 48 (mean age = 69.9 years, range 60 – 82 years) with sensorineural hearing loss, defined as 15 – 30 dB HL at 500 Hz, 20 – 40 dB HL at 1000 Hz, pure-tone average of ≤ 40 dB HL at 500, 1000 and 2000 Hz, word-recognition scores in quiet ≥ 40%, and no history of middle ear or retrocochlear pathology; age was restricted to 60 – 85 years	SPRINT Included 200 words, divided into 4 lists, spoken by a male speaker, mixed with multitalker babble (6 speakers) at 9 dB S/N	WIN Included 70 words spoken by a female speaker, divided into two 35-word lists that were administered sequentially The level of multitalker babble was fixed at 60 dB HL and the signal-to-babble ratio was varied by 4 dB increments	• To evaluate the concurrent validity of the SPRINT compared with the WIN in terms of recognition performances

Table A1: Characteristics of Included Clinical Studies					
First Author, Publication Year, Country	**Study Design**	**Patient Characteristics**	**Intervention(s)**	**Comparator(s)**	**Clinical Outcomes**
Wilson, 2007, USA[7]	Observational, repeated measures design	n = 96, divided into: n = 24 (18 women, 6 men; mean age = 24.9 years, SD = 2.8 years) with normal hearing, defined as ≤ 20 dB HL (mean pure-tone thresholds at the 250 to 8000 Hz octave frequencies were 8.1, 4.4, 2.7, 2.7, 4.6, and 2.3 dB HL) n = 72 (72 men; mean age = 70.8 years, SD = 9.5 years) with sensorineural hearing loss, defined as pure-tone measurements of 20 – 60 dB HL at 500, 1000, and 2000 Hz (mean at 500, 1000, and 2000 was 34.5 dB HL[SD = 8.0 dB HL]) Participants had not previously participated in listening experiments	HINT (Version 2.0; lists 1 and 8 were used), administered based on the HINT User's Manual Sentences were administered starting at 90 or 100 dB SPL (based on the participant's pure-tone audiogram result) and increased by 4 dB until the participant could correctly repeat the sentence Noise was presented at 90 dB SPL	WIN (Lists 1A and 2A were used), administered based on the WIN User's Manual Multitalker babble was fixed at 80 dB SPL; level of speech varied from 80 dB SPL – 104 dB SPL in 4 dB increments	• To evaluate the within- and between-group differences obtained on HINT and WIN. • Comparison between WIN and HINT using bivariate plots

dB = decibels; HINT = Hearing in Noise Test; HL = hearing loss; Hz = hertz; S/B = signal-to-babble ratio; SD = standard deviation; SPL = sound pressure level; SPRINT = Speech Recognition in Noise Test; WIN = Words in Noise Test

APPENDIX 4: Critical Appraisal of Included Publications

Table A2: Strengths and Limitations of Observational Studies using The Downs and Black Checklist[3]	
Strengths	**Limitations**
Vermiglio[4]	
• The study objectives and hypotheses were clearly described • Tests and administration processes were clearly described • Study outcomes were clearly described • No losses to follow up	• Investigators were not blinded to tests administered • A sample size calculation was not performed, therefore it is unclear whether type 2 error was present • The study population was not clearly described (unsure proportion of males and females) • Individuals with conductive hearing loss, non-response on the audiogram, asymmetrical hearing losses, and those with > +/- 2 standard deviations around the group mean for the pure-tone threshold groups were excluded from the study • Although statistical significance was reported, clinical significance of the results was not reported • Unclear whether the results would be generalizable to a typical auditory fitness for duty population
Wilson[5]	
• The study objectives were clearly described • Tests and administration processes were clearly described	• Investigators were not blinded to tests administered – potential issue given that the first author created the WIN • The study population was not clearly described (unsure proportion of males and females) • A total of 139 individuals did not undergo the WIN • Individuals with hearing loss defined as a pure-tone average of > 60 dB HL were excluded from the study • Study outcomes were not clearly described • Statistical analysis procedures were not included in the methods • Authors used means and standard deviations to report continuous variables, however, it is unclear whether these variables were normally distributed • Unclear whether the results would be generalizable to a typical auditory fitness for duty population
Wilson[6]	
• The study objectives were clearly described • Tests and administration processes were clearly described • No losses to follow up	• Inclusion of individuals with hearing loss was restricted to those 60 – 85 years old, with specific pure-tone thresholds and word-recognition scores in quiet, without middle ear or retrocochlear pathology

Table A2: Strengths and Limitations of Observational Studies using The Downs and Black Checklist[3]	
Strengths	**Limitations**
	• The study population was not clearly described (unsure proportion of males and females) • Investigators were not blinded to tests administered – potential issue given that the first author created the WIN • Study outcomes were not clearly described • Statistical analysis procedures were not included in the methods • Unclear whether the results would be generalizable to a typical auditory fitness for duty population
Wilson[7]	
• Study objective was clearly described • Study population was clearly described • Tests and administration processes were clearly described • Study outcomes were clearly described • No losses to follow up	• Investigators were not blinded to tests administered – potential issue given that the first author created the WIN • Different speakers were used for the different tests • Individuals with hearing loss > 60 dB HL at 500, 1000, and 2000 Hz were excluded from the study • Authors used means and standard deviations to report continuous variables, however, it is unclear whether these variables were normally distributed • Unclear whether the outcome measurement (mean 50% point dB S/B) translates into differences in clinical outcomes such as job performance • Unclear whether the results would be generalizable to a typical auditory fitness for duty population

dB = decibels; S/B = signal-to-babble ratio; WIN = Words in Noise Test

APPENDIX 5: MAIN STUDY FINDINGS AND AUTHOR'S CONCLUSIONS

Table A3: Summary of Findings of Included Studies	
Main Study Findings	**Author's Conclusions**
Vermiglio[4]	
Scatterplots of the HINT Composite Scores stratified by pure-tone threshold demonstrated similar performance across pure-tone threshold groups HINT Composite Scores – difference between those with normal hearing: • Slight = 1.06 (not significant) • Mild = 0.07 (not significant) • Moderate = 0.16 (not significant) • Severe = 0.38 (not significant) • Profound = 1.89 ($p < 0.05$)	• "For the Composite HINT scores, the only significant difference was found between the normal and profound pure-tone threshold groups. No significant differences were found between the normal group and the slight, mild, moderate or severe high-frequency pure-tone threshold groups." – page 783 • "Results from the present study are consistent with investigations that have shown a poor relationship between pure-tone thresholds and speech recognition in noise performance" – page 784 • "The audiogram is not useful to infer the ability to recognize speech in noise." – page 786
Wilson[5]	
• WIN 50% correct point in the left ear, mean = 13.8 dB SNR (SD = 5.1 dB) • WIN 50% correct point in the right ear, mean = 13.3 dB SNR (SD = 5.1 dB) • Those with a normal performance on the WIN, defined as a 50% correct point of ≤ 6 dB SNR, had a mean high-frequency pure-tone average of 19.7 dB HL (SD = 9.6 dB) in the left ear (n = 224, 6.8%) • Those with a normal performance on the WIN, defined as a 50% correct point of ≤ 6 dB SNR, had a mean high-frequency pure-tone average of 19.6 dB HL (SD = 9.9 dB) in the right ear (n = 265, 8.1%) • Bivariate plots of high-frequency pure-tone thresholds and the WIN demonstrated a clear linear relationship between the WIN 50% correct point and the pure-tone threshold average, with a slope of 2.5 dB/dB ($R^2 = 0.56$)	• "…as the hearing levels of the pure-tone thresholds increase, there is a corresponding increase in both age and the 50% point on the WIN." – page 414 • WIN performance was significantly better in the right ear compared to the left ear, which was likely reflective of administering the test first in the left ear • "Performance on the WIN was more closely associated with pure-tone thresholds than with age." – page 419 • "The results form the current WIN clinical study coupled with the results from the 12 laboratory studies on the WIN amply demonstrate the WIN (1) provides a valid and reliable measure of word recognition in background noise (babble), (2) enables word recognition both in quiet and in noise to be evaluated with the same works spoken by the same speaker, (3) can be used on a diverse clinic population from children through adults, and (4) is ready for routine clinical implementation." – page 419
Wilson[6]	
• SPRINT mean correct in participants with normal hearing was 92.5% (SD = 2.4%) • SPRINT mean correct in participants with hearing loss was 65.3% (SD = 11.2%) • WIN 50% point for people with normal hearing was 1.8 dB S/N • WIN 50% pint for people with hearing loss was 11.3 dB S/N • 65% correct was achieved at 9.0 dB S/N on the SPRINT, and 13.6 dB S/N on the WIN	• "On the SPRINT, 65% correct was achieved at 9.0 dB S/N, whereas on the WIN the 65% point was achieved at 13.6 dB S/N (calculated from the polynomial equation), which is a 4.6 dB difference. Based on data obtained in quiet in earlier investigations, this 4.6 dB difference is attributable to speaker differences between the two recordings used by the SPRINT and WIN paradigms." – page 553 • "A Pearson product-moment correlation indicated a significant relationship (r = -0.81, p

Table A3: Summary of Findings of Included Studies

Main Study Findings	Author's Conclusions
• Bivariate plot comparisons of the SPRINT and WIN demonstrated that as performances improved on the SPRINT, they also improved on the WIN • Pearson product-moment correlation for the performance between the SPRINT and the WIN was r = -0.81, p < 0.01	< 0.01) between the performances on the SPRINT and WIN. This highly related correlation indicates good concurrent validity of the SPRINT…" – page 553 • "Graphically and numerically the SPRINT and WIN were highly related, which is indicative of good concurrent validity of the SPRINT." – page 555
Wilson[7]	
• HINT 50% points for participants with normal hearing: 4.1 dB S/B (SD = 2.8) for list 1; 2.6 dB S/B (SD = 2.2) for list 8, 3.3 dB S/B (SD = 2.6) for both • HINT 50% points for participants with hearing loss: 9.0 dB S/B (SD = 4.9) for list 1; 8.9 dB S/B (SD = 4.6) for list 8, 8.9 dB S/B (SD = 4.7) for both • WIN 50% points for participants with normal hearing: 3.6 dB S/B (SD = 1.5) for list 1; 4.2 dB S/B (SD = 1.5) for list 2, 3.9 dB S/B (SD = 1.0) for both • WIN 50% points for participants with hearing loss: 13.9 dB S/B (SD = 4.5) for list 1; 14.2 dB S/B (SD = 4.4) for list 2, 14.0 dB S/B (SD = 4.73 for both • Bivariate plot comparison of the HINT and WIN demonstrated that almost all of the 50% points for the WIN had a better signal-to-noise ratio compared with the HINT 50% points, indicating that the WIN was a more sensitive measure for speech-in-noise compared to the HINT	• "This finding indicates for the listeners with hearing loss that as contextual cues in the speech material are reduced (e.g., going from HINT to WIN materials) and as recognition is more dependent upon acoustic cues, recognition performance decreases." – page 850 • "As expected, better recognition performance was obtained from the listeners with normal hearing than from the listeners with hearing loss." – page 855 • The authors concluded that the WIN provided "more separation in terms of recognition performance between the two groups of listeners" compared to the HINT materials, and that the "WIN be incorporated into routine clinic protocols as a speech-in-noise task" – page 855

dB = decibels; HINT = Hearing in Noise Test; S/B = signal-to-babble ratio; SD = standard deviation; S/N = signal-to-noise ratio; SNR = signal-to-noise ratio; SPRINT = Speech Recognition in Noise Test; WIN = Words in Noise Test